Technical Manual
for
S-Curve Tool
Version 1.0

(as of 09/12/12)

Points of Contact:
Bruce Parker, Bruce.M.Parker@navy.mil, (703) 604-3459
Brian Flynn, Brian.Flynn.ctr@navy.mil, (703) 692-4899
Richard Lee, RLee@technomics.net, (571) 366-1447
Peter Braxton, PBraxton@technomics.net, (571) 366-1431

Acknowledgements

Table of Contents

List of Tables

List of Figures

Introduction

The purpose of this document is to display/explain the mathematics behind the S-Curve Tool v1.0. Refer to the User Guide for additional details on the different tabs in the tool. Figure 1 shows a flowchart diagram of the S-Curve Tool v1.0. For the estimate(s), the user chooses Empirical (i.e., a set of outcomes from a Monte Carlo risk run), Parametric (e.g., enhanced Scenario-Based Method (eSBM) or parameters from an external risk analysis), or a Point Estimate (i.e., risk analysis not yet done). Historical adjustments are based on the Naval Center for Cost Analysis's (NCCA's) analysis of Selected Acquisition Reports (SARs) and are dependent on five different inputs: (1) commodity, (2) life cycle phase, (3) milestone, (4) inflation, and (5) quantity. If users decide not to apply historical adjustments to the estimate, they can proceed with the base s-curve that was generated.
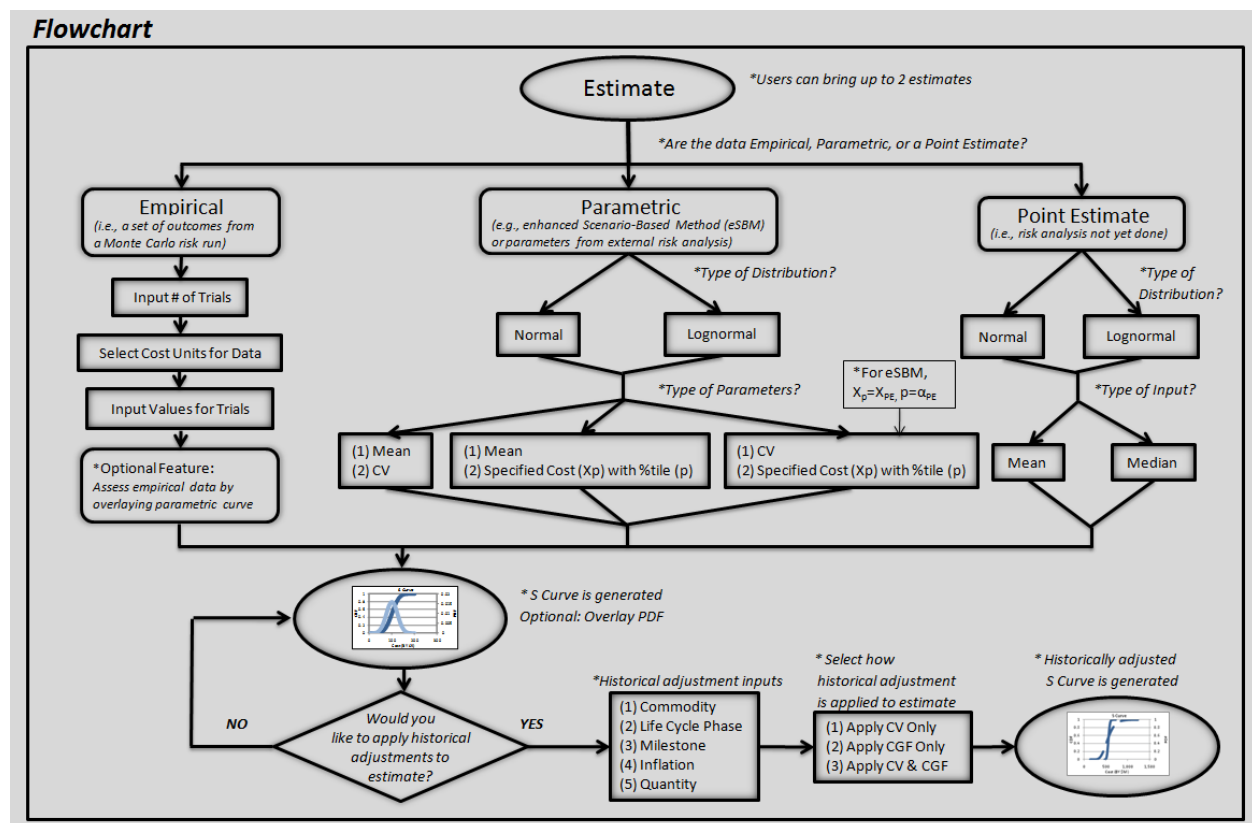
Figure 1: Flowchart Diagram of S-Curve Tool v1.0

Table 1 displays all the possible scenarios for defining an s-curve in the S-Curve Tool v1.0. With the exception of the Point Estimate, Lognormal, Median case (Scenario 11), all scenarios retain the value of the mean throughout the s-curve; scenario 11 retains the value of the median for the s-curve.

# of Scenarios	Estimate Type	Distribution	Type of Input
1	Empirical	-	-
2	Parametric	Normal	Mean and CV
3	Parametric	Normal	Mean and Specified Cost (w/ %tile)
4	Parametric	Normal	CV and Specified Cost (w/ %tile)
5	Parametric	Lognormal	Mean and CV
6	Parametric	Lognormal	Mean and Specified Cost (w/ %tile)
7	Parametric	Lognormal	CV and Specified Cost (w/ %tile)
8	Point Estimate	Normal	Mean
9	Point Estimate	Normal	Median
10	Point Estimate	Lognormal	Mean
11	Point Estimate	Lognormal	Median

Table 1: All Possible Scenarios for Defining an S-Curve in the S-Curve Tool v1.0

Empirical Estimate Type

This chapter provides further details for the Empirical estimate type. The S-Curve Tool follows the list of procedures to display the PDFs and CDFs of the empirical data.

1. Insert the number of trials and all Monte Carlo simulation outputs into the tool. The tool then automatically converts the cost units of the empirical (raw) data to the cost units selected in the "Inputs" tab using the conversion factors listed in Table 2.

Units for Empirical (Raw) Data	Converted to	Conversion Factor
$	$K	0.001
$K	$K	1
$M	$K	1,000
$B	$K	1,000,000
$	$M	0.000001
$K	$M	0.001
$M	$M	1
$B	$M	1,000
$	$B	0.000000001
$K	$B	0.000001
$M	$B	0.001
$B	$B	1

Table 2: Conversion Factors for Empirical (Raw) Data

2. The tool automatically ranks and sorts the entered values.
3. Using Equation 1 the tool automatically calculates the CDF and applies it to the sorted values. Calculations are dependent on the total # of trials that the user specified on the Inputs tab.

$$CDF_i = \frac{i}{total \ \# \ of \ trials} \tag{1}$$

4. Table 3 defines parameters that are derived from the Empirical (raw) data, that is from the Monte Carlo (or other statistical) output.

Derived Parameters From Empirical (raw) Data	Equation
Mean	$\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$
Standard Deviation	$s = \sqrt{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Coefficient of Variation (CV)	$\dfrac{s}{\bar{x}}$
Variance	s^2
Median	Middle value that separates the upper half from the lower half of the data set

Technomics

Mode	The most frequent value in the data set
Underlying Mean (Calculated for Lognormal distributions only)	$ln\left(\dfrac{\overline{x}}{\sqrt{1+CV^2}}\right)$
Underlying Standard Deviation (StDev) (Calculated for Lognormal distributions only)	$\sqrt{\ln\left(1+CV^2\right)}$
Min	Minimum value empirical raw data set
Max	Maximum value of empirical raw data set
Z_{min}	$\dfrac{x_{min}-\overline{x}}{s}$
Z_{max}	$\dfrac{x_{max}-\overline{x}}{s}$
Interval ($n_{smoothing}$ = 201 points for current version of the tool, refer to Table 4)	$\dfrac{Z_{max}-Z_{min}}{n_{smoothing}-1}$

**Table 3:** Parameters Derived From Empirical (Raw) Data

5. The tool incorporates a "smoothing" technique for the empirical data. This technique is only used to facilitate the display of data, NOT to re-calculate the derived parameters from the empirical (raw) data (shown in Table 3). The concept behind the "smoothing" technique is simply to constrain the display of the CDF and PDF to 201 points (200 intervals). By selecting 201 points, the PDF of the raw data can be easily recognized and the tool can run faster, since it doesn't have to store and perform calculations for 10,000 data points (maximum number of trials allowed for the tool). Table 4 shows the equations/descriptions for calculating the 201 points in the "smoothing" technique.

Calculated Values	Equation/Description
Z_i	From derived calculations, spread Z_{min} and Z_{max} into *equal intervals* of 201 points
X_i	$\overline{x}+Z_i*s$
X (smoothed)$_i$	From the equally spaced intervals, look up (approximate) the closest value to the raw data
Cumulative Distribution Function (CDF)	From the X(smoothed) column, look up the exact value of the CDF in the raw data column
Probability Density Function (PDF)	$\dfrac{\Delta CDF}{\Delta x\ (smoothed)}$

Table 4: Descriptions for Calculating the 201 Points in the "Smoothing" Technique

Shown below are two examples of the "smoothing" technique. As shown in Example 1 and Example 2, the "smoothing" technique is mainly used to "smooth" the PDF. There is no drastic change to the CDF.

Example 1: Number of trials for empirical (raw) data is GREATER THAN number of points for "smoothing" technique. In this case, 201 points from "smoothing" technique vs 10,000 Monte Carlo trials from Empirical (raw) data.
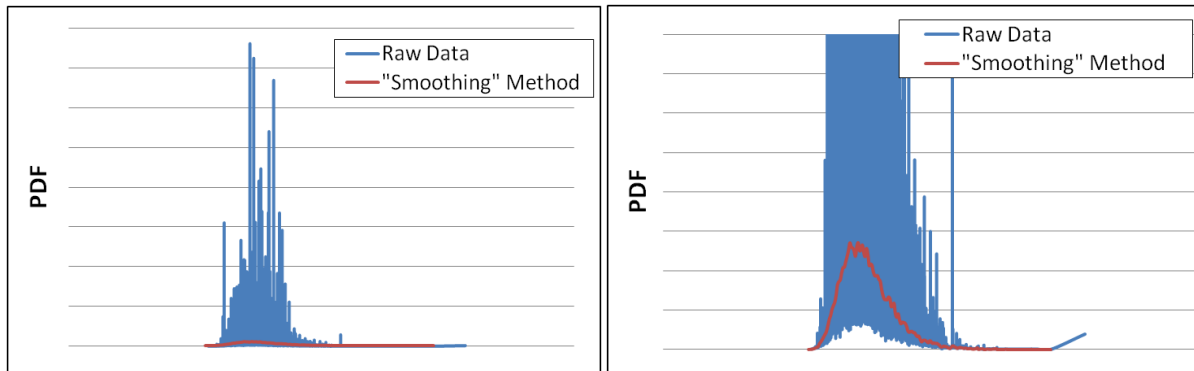
2a. Auto scale y-axis   **2b. Constrained y-axis**

Figure 2: Comparison between PDF of Empirical Data and "Smoothing" Technique for trials greater than 201 points
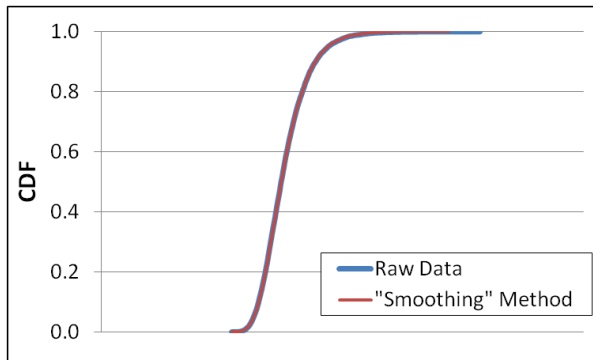
Figure 3: Comparison between CDF of Empirical Data and "Smoothing" Technique for trials greater than 201 points

Example 2: Number of trials for empirical (raw) data is LESS THAN number of points for "smoothing" technique. In this case, 201 points from "smoothing" technique vs 100 Monte Carlo trials from Empirical (raw) data.
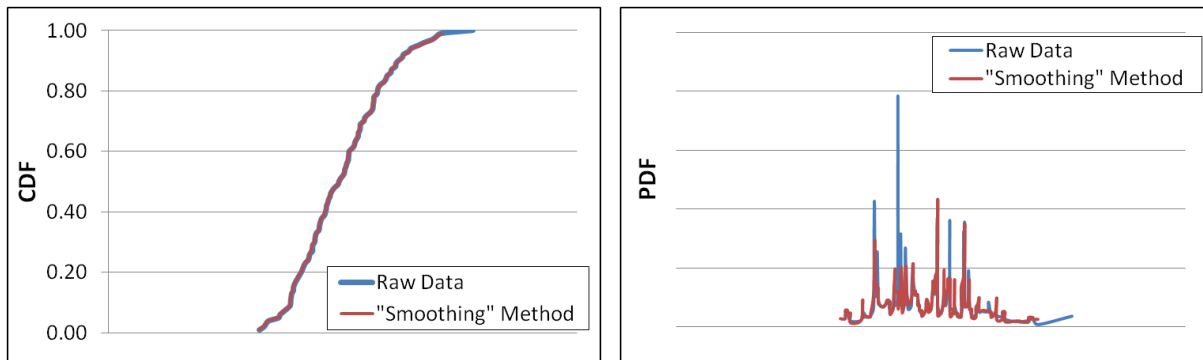
4a. Comparison of CDF   **4b. Comparison of PDF**

Figure 4: Comparison between Empirical Data and "Smoothing" Technique for trials less than 201 points

6. As described in the User Manual, the tool allows the user to make adjustments based on historical data. Using the parameters derived from the empirical (raw) data in Table 3 and the equations/descriptions for the calculations of the 201 points in Table 4, Equation 2 can be used to make these adjustments. This again facilitates a quicker process.

$$x_{ha_i} = CGF\left[\bar{x} + \frac{CV_{hist}}{CV} * (x_i - \bar{x})\right] \qquad (2)$$

where x_{ha_i} is the historically adjusted x value.

Parametric and Point Estimate Types

This chapter provides further details for the Parametric and Point Estimate types. As previously stated, the value of the mean is retained for all estimate types except for the Point Estimate, Lognormal, Median case, which retains the value of the median. The mean is equal to the median for Normal distributions, and therefore, the Point Estimate, Normal, Mean case retains the mean.

Table 5 displays the equations for all types of inputs for the Parametric and Point Estimate types (refer to scenarios 2 to 11 in Table 1). The equations in Table 5 are also categorized by Normal and Lognormal distributions. Cells that are filled in green contain equations related to the user inputs in the S-Curve Tool, while all other cells that are not filled contain equations used to calculate the derived parameters from the user inputs. The Point Estimate type can be viewed as a unique case of the Parametric Estimate type.

The Parametric Estimate with mean and CV inputs (scenario 2 - Normal distribution and scenario 5 - Lognormal distribution in Table 1) are comparable to the Point Estimate (Mean) distributions (scenario 8 - Normal distribution and scenario 10 – Lognormal distribution in Table 1). These comparable scenarios are highlighted in the header of Table 5 with blue filled cells. For the Point Estimate (Mean) case, the tool applies a CV of 0.0001 to the equations that are listed for the Parametric estimate with mean and CV inputs.

The Point Estimate with CV and specified cost (X_p, p) inputs (scenario 4 - Normal distribution and scenario 7 - Lognormal distribution in Table 1) are comparable to the Point Estimate (Median) distributions (scenario 9 - Normal distribution and scenario 11 – Lognormal distribution in Table 1). These comparable scenarios are highlighted in the header of Table 5 with orange filled cells. For the Point Estimate (Median) case, the tool applies a CV of 0.0001 and a percent (p) of 50% to the equations that are listed for the Parametric Estimate with CV and specified cost inputs.

$Z_p = \Phi^{-1}(p)$		**Parametric**			**Point Estimate**	
		Mean and CV	Mean and Specified Cost (X_p, p)	CV and Specified Cost (X_p, p)	Mean (assume CV=0.0001)	Median (assume CV=0.0001, p=0.5)
Normal Distribution	Mean (μ)	μ	μ	$\dfrac{X_p}{1+Z_p \cdot CV}$	μ	$\dfrac{X_p}{1+Z_p \cdot CV}$
	CV (σ/μ)	CV	$\dfrac{\left(\dfrac{X_p}{\mu}-1\right)}{Z_p}$	CV	CV	$\dfrac{X_p \cdot CV}{1+Z_p \cdot CV}$
	Std Dev (σ)	$CV \cdot \mu$	$\dfrac{X_p - \mu}{Z_p}$	$\dfrac{X_p \cdot CV}{1+Z_p \cdot CV}$	$CV \cdot \mu$	CV
	Variance (σ^2)	σ^2	σ^2	σ^2	σ^2	σ^2
	Median (= μ)	μ	μ	μ	μ	μ
	Mode (= μ)	μ	μ	μ	μ	μ
Lognormal Distribution	Mean (E[X])	$E(X)$	$E(X)$	$e^{\mu+\frac{\sigma^2}{2}}$	$E(X)$	$e^{\mu+\frac{\sigma^2}{2}}$
	CV (s. d.(X) / E[X])	CV	$\sqrt{e^{\sigma^2}-1}$	CV	CV	CV
	Std Dev (s.d.[X])	$CV \cdot E(X)$	$CV \cdot E(X)$	$CV \cdot E(X)$	$CV \cdot E(X)$	$CV \cdot E(X)$
	Variance (Var[X])	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
	Median (e^μ)	e^μ	e^μ	e^μ	e^μ	e^μ
	Mode (e^(μ-σ^2))	$e^{\mu-\sigma^2}$	$e^{\mu-\sigma^2}$	$e^{\mu-\sigma^2}$	$e^{\mu-\sigma^2}$	$e^{\mu-\sigma^2}$
	Mean Underlying Normal (μ)	$\ln\left(\dfrac{E(X)}{\sqrt{1+CV^2}}\right)$	$\ln\left(\dfrac{E(X)}{\sqrt{1+CV^2}}\right)$	$\ln X_p - Z_p\sqrt{\ln(1+CV^2)}$	$\ln\left(\dfrac{E(X)}{\sqrt{1+CV^2}}\right)$	$\ln X_p - Z_p\sqrt{\ln(1+CV^2)}$
	St. Dev. Underlying Normal (σ)	$\sqrt{\ln(1+CV^2)}$	$\dfrac{Z_p \pm}{\sqrt{Z_p^2 - 2\ln\left(\dfrac{X_p}{E(X)}\right)}}$	$\sqrt{\ln(1+CV^2)}$	$\sqrt{\ln(1+CV^2)}$	$\sqrt{\ln(1+CV^2)}$

Table 5: Derivations for the Parametric and Point Estimate Types

Historical Adjustments

The historical coefficients of variations (CVs) and cost growth factors (CGFs) are based on the Naval Center for Cost Analysis's (NCCA's) analysis of Selected Acquisition Reports (SARs) and are dependent on five different inputs: (1) commodity, (2) life cycle phase, (3) milestone, (4) inflation, and (5) quantity. The first step in NCCA's calculation of CVs and CGFs was to extract the original estimate and the current estimate for a given program from the SAR Summary Sheets (both in base year and then year dollars). From the extracted data, the CGF was calculated by dividing the current estimate by the original estimate, shown in Equation 3.

$$CGF = \frac{Current\ Estimate}{Original\ Estimate} \tag{3}$$

NCCA's approach in deriving coefficients of variations (CVs) was to categorize the CGFs of all programs by the five historical adjustment inputs. The mean and the standard deviation were calculated from the filtered data set of CGFs.

$$CV = \frac{standard\ deviaton}{mean} \tag{4}$$

Note that the calculated CV is the CV of CGFs, NOT the CV of cost.

After the historical CVs and CGFs are obtained, users select one of three options to apply the historical adjustment to the estimate: (1) apply CV only ("flattening" the s-curve), (2) apply CGF only ("shifting" the s-curve, or (3) apply CV & CGF ("flattening" and "shifting" the s-curve). If users decide not to apply historically adjustments to the estimate, they can proceed with the base s-curve that was generated from the tool. Table 6 shows how the historical CVs and/or historical CGFs are applied to the base s-curve. As stated in the previous sections, the Point Estimate (Median) case (refer to scenario 11 in Table 1) is the only scenario that retains the median. Therefore, historical adjustments that are applied to this scenario are treated differently from all other cases (shown in the bottom of Table 6). For example, if the user selects "CV Only" for scenario 11 in Table 6, the historically adjusted s-curve "pivots" on the median, whereas for all other scenarios, the historically adjusted s-curve "pivots" on the mean.

	CV Only	CGF Only	CV & CGF Only
For ALL cases except *Special Case: Point Estimate, Lognormal, Median			
Hist. Adj. Mean (Mean$_{HA}$)	Base Mean	Base Mean*Suggested CGF	Base Mean*Suggested CGF
Hist. Adj. CV (CV$_{HA}$)	Suggested CV	Base CV	Suggested CV
Hist. Adj. StDev (StDev$_{HA}$)	Base Mean*Base CV	Base Mean*Base CV	Base Mean*Base CV
Underlying Mean (only calculated if Lognormal is selected)	$\ln\left(\dfrac{Mean_{HA}}{\sqrt{1+CV_{HA}^{2}}}\right)$	$\ln\left(\dfrac{Mean_{HA}}{\sqrt{1+CV_{HA}^{2}}}\right)$	$\ln\left(\dfrac{Mean_{HA}}{\sqrt{1+CV_{HA}^{2}}}\right)$
Underlying StDev (only calculated if Lognormal is selected)	$\sqrt{\ln\left(1+CV_{HA}^{2}\right)}$	$\sqrt{\ln\left(1+CV_{HA}^{2}\right)}$	$\sqrt{\ln\left(1+CV_{HA}^{2}\right)}$
Hist. Adj. Median (Median$_{HA}$)	Normal: Mean$_{HA}$ / Lognormal: $e^{Mean_{HA}}$	Normal: Mean$_{HA}$ / Lognormal: $e^{Mean_{HA}}$	Normal: Mean$_{HA}$ / Lognormal: $e^{Mean_{HA}}$
*Special Case: Point Estimate, Lognormal, Median			
Hist. Adj. Mean (Mean$_{HA}$)	$e^{\mu_{ha}+\frac{\sigma_{ha}^{2}}{2}}$	$e^{\mu_{ha}+\frac{\sigma_{ha}^{2}}{2}}$	$e^{\mu_{ha}+\frac{\sigma_{ha}^{2}}{2}}$
Hist. Adj. CV (CV$_{HA}$)	Same as above ("for all cases")		
Hist. Adj. StDev (StDev$_{HA}$)	Same as above ("for all cases")		
Underlying Mean (μ$_{ha}$)	$\ln X_{p}$	$\ln X_{p}$	$\ln X_{p}$
Underlying StDev (σ$_{ha}$)	Same as above ("for all cases")		
Hist. Adj. Median (Median$_{HA}$)	Base Median	Base Median*Suggested CGF	Base Median*Suggested CGF

Table 6: Applying CVs and/or CGFs to S-Curves Using S-Curve Tool

Chart Options

This chapter provides further details on the derived parameters and all calculated points for the chart options in the tool. These calculations are mainly used in the "Benchmarking" and "Reconciliation" tabs. The parameters that can be applied to a chart are (1) Mean, (2) Median, (3) Custom Cost, (4) 20th Percentile, (5) 80th Percentile, and (6) Custom Percentile. For the "Custom Cost" selection, users input a cost of their choice and the tool calculates the percentile. Similarly, for the "Custom Percentile" selection, users input a percentile of their choice and the tool calculates the cost. The equations/approaches used to calculate the chart parameters for Empirical data, data with a Normal distribution (for both the Parametric and Point Estimate cases), and data with Lognormal distribution (for both the Parametric and Point Estimate cases) are listed below.

1. Empirical data (no distribution)

    Calculate all values from raw Empirical data.

2. Equation 5 shows the Normal Distribution PDF for both the Parametric and Point Estimate cases.

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

3. Equation 6 shows Lognormal Distribution PDF for both the Parametric and Point Estimate cases.

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \tag{6}$$

Further details are provided in Table 7, which displays equations for Z, cost (X), PDF, and CDF. These equations apply to all 201 points used to create the s-curve.

Estimate Type	Distribution	Z_i	Cost $(x)_i$	PDF_i	CDF_i
Base Estimate					
Empirical	-	*blank*	$CGF\left[\bar{x} + \dfrac{CV_{hist}}{CV} * (x_i - \bar{x})\right]$	$\dfrac{\Delta CDF}{\Delta x}$	$CDF_i = \dfrac{i}{200}$
Parametric Point Estimate	Normal	-4 to 4 in equal intervals	$\mu + x_i * \sigma$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$	$\dfrac{1}{2}\left[1 + erf\left(\dfrac{x_i - \mu}{\sqrt{2\sigma^2}}\right)\right]$
Parametric Point Estimate	Lognormal	-3 to 6 in equal intervals	$E(X) + x_i * \sigma$	$\dfrac{1}{x_i\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x_i-\mu)^2}{2\sigma^2}}$	$\dfrac{1}{2} + \dfrac{1}{2}\left[1 + erf\left(\dfrac{\ln x_i - \mu}{\sqrt{2\sigma^2}}\right)\right]$
Historical Adjustment (if checkbox is checked, then calculations below will show, otherwise, spit out "#N/A")					
Empirical	-	*Equal to Base Estimate*	$CGF\left[\bar{x}_{HA} + \dfrac{CV_{hist}}{CV} * (x_i - \bar{x}_{HA})\right]$	$\dfrac{\Delta CDF}{\Delta x}$	$CDF_i = \dfrac{i}{200}$
Parametric Point Estimate	Normal	*Equal to Base Estimate*	$\mu_{HA} + x_i * \sigma_{HA}$	$\dfrac{1}{\sqrt{2\pi\sigma_{HA}^2}}e^{-\frac{(x_i-\mu_{HA})^2}{2\sigma_{HA}^2}}$	$\dfrac{1}{2}\left[1 + erf\left(\dfrac{x_i - \mu_{HA}}{\sqrt{2\sigma_{HA}^2}}\right)\right]$
Parametric Point Estimate	Lognormal		$E(X)_{HA} + x_i * \sigma_{HA}$	$\dfrac{1}{x_i\sqrt{2\pi\sigma_{HA}^2}}e^{-\frac{(\ln x_i-\mu_{HA})^2}{2\sigma_{HA}^2}}$	$\dfrac{1}{2} + \dfrac{1}{2}\left[1 + erf\left(\dfrac{\ln x_i - \mu_{HA}}{\sqrt{2\sigma_{HA}^2}}\right)\right]$

Table 7: Calculations for All Displayed Data Points on Charts